BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 5: e75554
doi: 10.3897/biss.5.75554

OPEN ACCESS

Conference Abstract

TaxonWorks: Character state matrices and identification tools

Dmitry A Dmitriev‡, Matthew Yoder‡

‡ University of Illinois at Urbana-Champaign, Champaign, United States of America

Corresponding author: Dmitry A Dmitriev (arboridia@gmail.com)

Received: 21 Sep 2021 | Published: 21 Sep 2021

Citation: Dmitriev DA, Yoder M (2021) TaxonWorks: Character state matrices and identification tools. Biodiversity Information Science and Standards 5: e75554. https://doi.org/10.3897/biss.5.75554

Abstract

TaxonWorks is an integrated web-based application for practicing taxonomists and biodiversity specialists. It is focused on promoting collaboration between researchers and developers. TaxonWorks has a modular structure that enables various components of the application to target specific needs and requirements of different groups of users. Specific areas of interest may include nomenclature-related tasks (Yoder and Dmitriev 2021) designed to help assemble and validate scientific name checklists of a target group of organisms; and collection management tasks, including interfaces to create, filter, and edit collecting events, collection objects, and loans. This presentation focuses on matrix-related tools integrated into TaxonWorks. A matrix, which could either be used for phylogenetic analysis or to build an identification key, is structured as a table where columns represent numerous characters that could be used to describe a set of entities, taxa or specimens (presented as rows of the table). Each cell of the table may contain observations for specific character/entity combinations. TaxonWorks does not generate a table for each a particular matrix—all observations are stored as graphs. This structure allows building of a matrix of an unlimited size as well as reuse of individual observations in multiple matrices. For matrix columns, TaxonWorks supports a variety of different kinds of characters or descriptors: qualitative, presence/absence, quantitative, sample, gene, free text, and media. Each character may have specific properties, for example a qualitative descriptor may have numerous characters states, and a quantitative descriptor may have a measurement unit defined. For an entity in a matrix row, TaxonWorks supports either collection objects (specimens) or taxa as Operational Taxonomic Units (OTU). OTUs could

© Dmitriev D, Yoder M. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

either be linked to nomenclature or be stand alone entities (e.g., representing undescribed species).

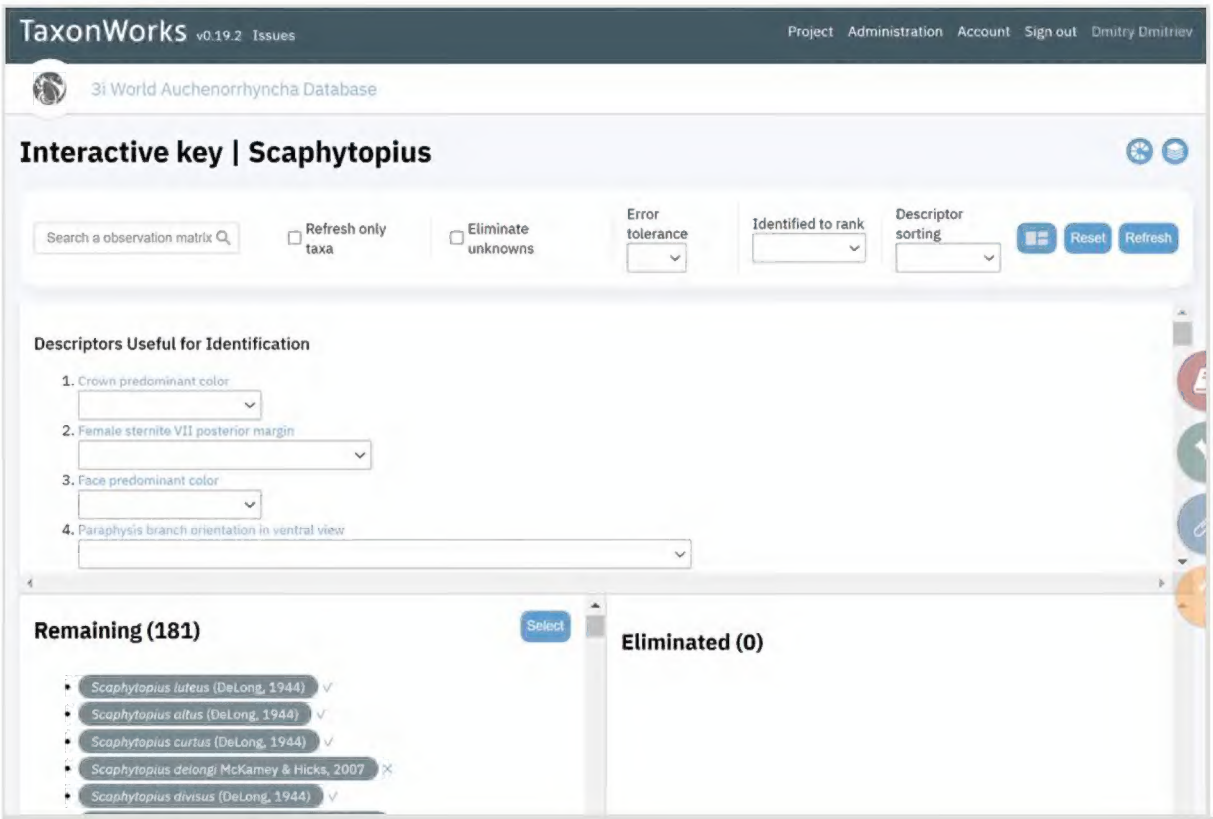

Figure 1.

Screenshot of the "Interactive key" interface in TaxonWorks from the "3i World Auchenorrhyncha database" (Dmitriev 2003).

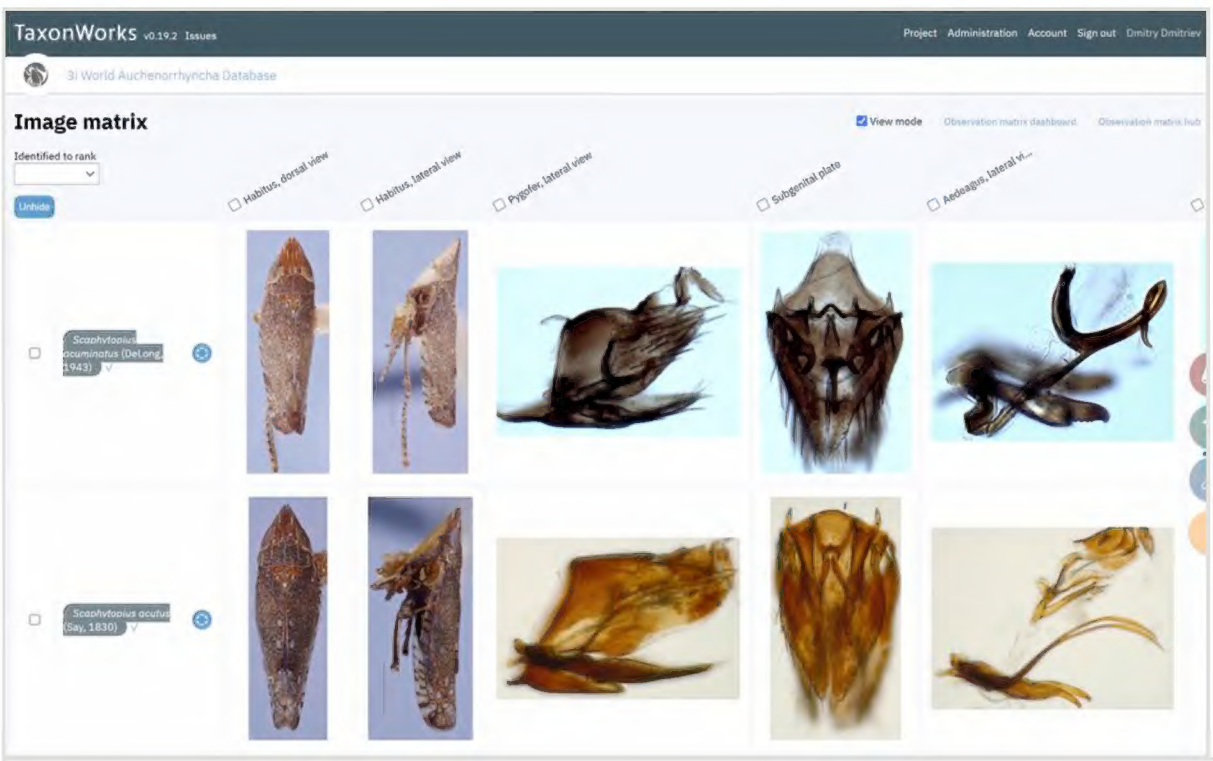

Figure 2.

Screenshot of the "Image matrix" interface in TaxonWorks showing the "3i World Auchenorrhyncha database" (Dmitriev 2003).

The matrix, once built, could serve several purposes. A matrix based on qualitative and quantitative characters could be used to build an interactive key (Fig. 1), construct

standardized natural language descriptions for each entity, and determine a diagnosis (a minimal set of characters that separate one entity from all others). It could also be exported and used for phylogenetic analysis or to build an interactive key in an external application. TaxonWorks supports export files in several formats, including Nexus, TNT, NeXML. Application Programming Interfaces (API) are also available. A matrix based on media descriptors could be used as a pictorial identification tool (Fig. 2).

Keywords

bioinformatics, cybertaxonomy, matrix, interactive identification key

Presenting author

Dmitry A. Dmitriev

Acknowledgements

This work is supported by the Species File Group and the U.S. National Science Foundation (NSF DEB 16-39601).

References

- Dmitriev D (2003) 3i World Auchenorrhyncha Database. http://dmitriev.speciesfile.org. Accessed on: 2021-9-12.
- Yoder M, Dmitriev D (2021) Nomenclature over 5 years in TaxonWorks: Approach, implementation, limitations and outcomes. Biodiversity Information Science and Standards 5: e75441. https://doi.org/10.3897/biss.5.75441